UNIVERSITAT DE BARCELONA

Estudios de Fonética Experimental
*Journal of Experimental Phonetics*
ISSN: 1575-5533 · ISSN-e: 2385-3573

EFE
Estudios de Fonética Experimental
Journal of Experimental Phonetics

# Sentence-level rhythmic regularity and accentual-group phrasing increase linguistic resemblance between Spanish-speaking interlocutors

Leonardo Francisco Barón-Birchenall [1] 0000-0002-4286-3424

[1] Corporación Universitaria Minuto de Dios – UNIMINUTO (Colombia)

DOI: 10.1344/efe-2025-34-129-150

Corresponding address: laescaladesol@gmail.com

Received: 28-07-2024 Accepted: 11-06-2025 Published: 23-09-2025

Barón-Birchenall, L. F. (2025). Sentence-level rhythmic regularity and accentual-group phrasing increase linguistic resemblance between Spanish-speaking interlocutors. *Estudios de Fonética Experimental*, *34*, 129–150. https://doi.org/10.1344/efe-2025-34-129-150



## Abstract

In two empirical studies, we examined the impact of speech rhythm on linguistic accommodation during conversational interactions. We compared dyadic interactions between Spanish speakers using regular and irregular rhythmic sentences structured by accentual feet (inter-stress intervals) and accentual groups. Acoustic-prosodic features (fundamental frequency, speech rate, response times) and resemblance measures (Euclidean distances) were analyzed. Additionally, independent listeners rated rhythmic resemblance between dyads. Linear mixed-effects models revealed greater resemblance in rhythm and F0 range, lower F0 mean, and narrower F0 range in interactions involving regular rhythmic sentences arranged in accentual groups. No differences were observed in response times or convergence measures between interlocutors across conditions. Some of our findings support the notion that rhythmic accommodation operates through a multilevel non-conscious mechanism, similar to that proposed by the interactive alignment model.

## Keywords

linguistic accommodation; speech rhythm; temporal regularity; accentual phrasing; shadowing task

## La regularitat rítmica a nivell oracional i el fraseig en grups accentuals augmenten la similitud lingüística entre parlants de castellà

RESUM

En dos estudis empírics, vam explorar l'impacte del ritme del discurs en l'acomodació lingüística durant les interaccions conversacionals. Vam comparar interaccions diàdiques entre parlants de castellà utilitzant frases rítmiques regulars i irregulars estructurades per peus accentuals (intervals entre accents) i grups accentuals. Es van analitzar característiques acústiques-prosòdiques (freqüència fonamental, velocitat de parla, temps de resposta) i mesures de similitud (distàncies euclidianes). A més, els oients van valorar la similitud rítmica entre díades. Els models lineals mixtos van revelar una major semblança en el ritme i el rang de F0, una mitjana d'F0 més baixa i un rang d'F0 més reduït durant les interaccions amb frases rítmiques regulars organitzades en grups accentuals. Alguns dels nostres resultats donen suport a la idea que l'acomodació rítmica opera mitjançant un mecanisme no conscient i multinivell, similar al proposat pel model d'alineació interactiva.

MOTS CLAU

acomodació lingüística; ritme del discurs; regularitat temporal; fraseig accentual; tasca de repetició immediata

## La regularidad rítmica a nivel oracional y el fraseo en grupos acentuales aumentan la similitud lingüística entre hablantes de español

RESUMEN

En dos estudios empíricos, exploramos el impacto del ritmo del discurso en la acomodación lingüística durante las interacciones conversacionales. Comparamos interacciones diádicas entre hablantes de español utilizando oraciones rítmicas regulares e irregulares estructuradas por pies acentuales (intervalos entre acentos) y grupos acentuales. Se analizaron características acústicas-prosódicas (frecuencia fundamental, velocidad del habla, tiempos de respuesta) y medidas de similitud (distancias euclidianas). Además, los oyentes evaluaron la similitud rítmica entre díadas. Los modelos lineales mixtos revelaron una mayor semejanza en el ritmo y en el rango de F0, una media de F0 más baja y un rango de F0 más reducido durante las interacciones con oraciones rítmicas regulares organizadas en grupos acentuales. Algunos de nuestros resultados respaldan la idea de que la acomodación rítmica opera mediante un mecanismo no consciente y multinivel, similar al propuesto por el modelo de alineación interactiva.

PALABRAS CLAVE

acomodación lingüística; ritmo del discurso; regularidad temporal; fraseo acentual; tarea de repetición inmediata

## 1. Introduction

The periodicity of auditory stimuli is crucial for sound perception and processing (Jones et al., 2002). While there is an ongoing debate in the field of linguistics about whether this periodicity is inherent in the acoustic speech signal or a perceptual construct (e.g., Brown et al., 2015), it is generally accepted that speech rhythm influences behavioral coordination, including linguistic interactions (e.g., Cummins, 2009). However, the exact mechanisms underlying this influence remain unclear. Moreover, specific studies examining the impact of speech rhythm on linguistic interactions among Spanish speakers are, to our knowledge, still lacking. The present study aims to contribute to this gap by systematically analyzing how rhythmic regularity and phonological phrasing shape linguistic accommodation in conversational settings.

As a starting point, we understood speech rhythm in terms of temporally regular iterations of stressed and unstressed syllables (as detailed in the next section). We believe that there are different effects of such temporal iterations on the outcomes of linguistic accommodation between interlocutors depending on the type of phrasing in which the regularity is implemented, specifically: accentual feet versus accentual groups arrangements. In the remaining of this section, we specify our stand on speech rhythm, detail the two types of phrasing employed in the experiments, and introduce the concept of linguistic accommodation.

### 1.1. Speech rhythm

For nearly 70 years, and more intensively during the past 30 years, linguists have been trying to establish what speech rhythm is, and how can it be quantified (Harris, 2015). Definition, validity, and underlying mechanisms of speech rhythm are subjects of an ongoing discussion. For instance, in the opinion of Arvaniti (2009), rhythm is not to be found in the acoustic speech signal or in the movements of the person who speaks. Rather, it is probably better understood as a link between the movement and the signal, a property of the act of speaking, and not of the speech signal itself.

On the other hand, as believed by Gibbon (2015), there is no such thing as speech rhythm, but speech rhythms, which consist of "temporally regular iterations of events which embody alternating strong and weak values of an observable parameter" (p. 115). In this definition, an observable parameter is a neutral concept between an auditory phenomenon and a set of acoustic factors. Alternating strong and weak values may also be referred to as loud-soft, stressed-unstressed, prominent-nonprominent, and even consonant-vowel.

Based on the aforementioned definition of speech rhythm, advanced by Gibbon (2015), the following restricted definition is employed in this paper: the rhythm of speech consists of temporal iterations of strong and weak values of lexical stress (i.e., stressed and unstressed syllables), which are constituted, in turn, by an unknown combination of duration, intensity, and pitch (further information about lexical stress implementation in Spanish can be found in Hualde, 2009). Note that this definition may not be applicable to tone or pitch-accent languages, in which a broader view of rhythm, such as a regular iteration of events, being either related to prominence or to constituent boundaries, would be more appropriate.

### 1.2. Phonological phrasing

In the experiments presented in this paper, we compare the two most studied forms of phonological phrasing in Spanish, both containing at least one syllable: *accentual feet*, which primarily depend on metric relations at the phonological level, and *accentual groups*, which also rely on syntactic connections between words.

Accentual feet are composed of a stressed syllable (or vowel) and all following unstressed segments up to the segment immediately preceding the next stressed syllable (Almeida, 1993, 1997; Cantero, 2002). The minimal foot in Spanish may be a monosyllable, as long as it is bimoraic (i.e., containing two moras; a mora is a time unit equivalent to a short vowel: *mes* 'month', and *sol* 'sun' are bimoraic syllables). Additionally, any monosyllabic content word forms a foot (Grau, 2013). For example (stressed syllables are represented by a number 1, unstressed syllables by a number 0, accentual feet are shown in brackets):

(1) Pedro lo sabe (*Peter knows it*)
[Pedro lo] [sabe]
[100] [10]

(2) Un perro gigante (*A giant dog*)
[Un] [perro gi] [gante]
[1] [100] [10]

However, in some studies on Spanish phonology, this particular type of phrasing is referred to as the inter-stress interval (Dauer, 1983; Borzone & Signorini, 1983). For instance, in Borzone and Signorini (1983, p. 119), "inter-stress intervals were measured from the consonant/vowel onset of the syllable bearing the stress to the consonant/vowel onset of the next syllable bearing stress." These inter-stress intervals, as presented in Dauer (1983) and Borzone and Signorini (1983), are interpreted by Toledo (2009, 2010) as accentual feet. Henceforth, we will refer to this type of phrasing as accentual feet.

Accentual groups in Spanish, on the other hand, have been studied by several researchers, who sometimes use different terms to describe them (e.g., Alcoba, 2007; Almeida, 1997; Mora et al., 1999). In this paper, we adopt the definition of accentual group provided by Hualde and Nadeu (2014), who refer to this type of phrasing principle as a phonetic group: a morphological word and any unstressed functional words preceding it. For example (accentual groups are shown in brackets):

(3) Amistad verdadera (*True friendship*)
[Amistad] [verdadera]
[001] [0010]

(4) Contra lo tratado durante la cita (*Against the agreement made during the appointment*)
[Contra lo tratado] [durante la cita]
[000010] [000010]

### 1.3. Linguistic accommodation

During conversational interactions such as tutoring, chatting with friends, and spoken negotiations, speakers' verbal behaviors undergo a series of changes due to the characteristics of their interlocutor and the nature of the interaction itself. This phenomenon, which is pervasively present in every social interaction, is referred to as linguistic accommodation. The process of accommodation is influenced by social, cultural, and personal aspects, including perceived social status, social biases, language background, perception of attractiveness, gender, and the desire to gain approval (e.g., Babel, 2011; Giles et al., 1991).

Linguistic accommodation has been studied with respect to several features, including speech rate and rhythm, pitch level and range, pitch accents, voice intensity and quality, vowels, pauses, and sentence duration, phonetic repertoire, regional accent, speaking and linguistic style, syntactic complexity, lexical choices, sub-phonemic features, and turn-taking (for an overview, see Barón, 2023; De Looze et al., 2014; and Pardo, 2006).

Research on linguistic accommodation has yielded mixed results, especially concerning phonetic accommodation. Acoustic analyses of the speech signal during conversations have proven to be inconsistent with perceptual judgments of the same interactions (Pardo, 2013; Barón, 2018, 2023). Moreover, despite convergence (i.e., a symmetric or asymmetric increase in similarity of diverse linguistic and paralinguistic features of two or more individuals during an interaction) being treated as the default form of accommodation in conversations, other types of accommodation, such as proximity (the maintenance of a certain degree of similarity during an interaction) or even divergence (an increasing difference between speakers' speech features and associated behaviors), may arise (Heath, 2015; see also De Looze & Rauzy, 2011, for further types of accommodation).

In this study, we use the term *resemblance* as a general synonym for similarity, without implying an increase or decrease in its degree. In cases where such a directional change is observed, the terms convergence or divergence are used, as previously explained.

There are two predominant theoretical models that deal with the characteristics of the cognitive-behavioral systems underlying linguistic accommodation: The Communication Accommodation Theory (CAT; Giles et al., 1991) and the Interactive Alignment Model (IAM; Pickering & Garrod, 2004). CAT originates from social psychology and sociolinguistics, emphasizing intentional-social factors and the adaptive benefits of accommodation. It suggests that adapting to interlocutors' communicative behavior fosters social approval and efficient communication, linking perceived behavioral similarity to positive attributions (Ruch et al., 2018).

In contrast, IAM originates from cognitive psychology and psycholinguistics, emphasizing mechanistic cognitive processes leading to accommodation. It posits that mutual understanding relies on adaptive processes across various linguistic levels (lexical, syntactic, semantic). Alignment at these levels contributes to aligning mental representations, essential for a successful conversation (Ruch et al., 2018; Xu & Reitter, 2016).

## 2. Experiments

### 2.1. Experiment 1: Acoustic task

#### 2.1.1. Participants

Twenty-four native Spanish speakers—12 females and 12 males—ranging from 18 to 28 years of age participated in the study. Participants were unacquainted before the experiment and had lived in Bogotá (Colombia) for the majority of their lives, ensuring no different dialects were involved in the experiment. To be part of the study, participants had to meet three conditions: (1) not present speech or hearing disorders; (2) not have learned a second language during childhood; and (3) not have formal musical training. The last condition was imposed because audio–motor training can influence the perception of rhythm and enhance listeners' sensitivity to metrical deviants (Cason et al., 2015). All participants were university students who received compensation in the form of extra course credit.

#### 2.1.2. Stimuli

For the purpose of this study, we created 64 Spanish sentences, each containing nine syllables and sharing the same syntactic structure: subject + verb + complement (see Appendix). Every sentence is composed of six words, all belonging to the 5,000 most frequent words in Spanish (according to Real Academia Española, 2008). Additionally, each sentence is structured with three accentual feet or accentual groups, each featuring a single strong syllable and without punctuation.

The sentences were equally divided into four blocks, with 16 sentences per block. Each block featured a specific rhythmic structure achieved through the arrangement of different types of words (oxytones, paroxytones, proparoxytones, and unstressed words) into regular or irregular accentual feet, and regular or irregular accentual groups. Regular groups and feet had a 3 / 3 / 3 syllabic distribution, while irregular groups and feet had a 2 / 4 / 3 syllabic distribution, as follows (stressed syllables are represented by a number 1, unstressed syllables by a number 0, and accentual feet or groups are shown in brackets):

a) Regular feet (RF)
   Mario te vio sin la máquina (Mario saw you without the machine)
   [100] [100] [100]
b) Regular groups (RG)
   La casa se vende por partes (The house is sold by parts)
   [La casa] [se vende] [por partes]
   [010] [010] [010]
c) Irregular feet (IF)
   Sol me cuenta de su sábado (*Sol tells me about her* Saturday)
   [Sol me] [cuenta de su] [sábado]
   [10] [1000] [100]

d) Irregular groups (IG)
Mario se nos quedó sin novia (*Mario ended up without a girlfriend*)
[Mario] [se nos quedó] [sin novia]
[10] [0001] [010]

In addition, we ensured the avoidance of all forms of resyllabification between word limits. Voiced stops, /b/, /d/, and /g/, were avoided in sentence initial positions to facilitate onset measurement, following the procedure outlined by Späth et al. (2016). Within each block of stimuli, neither content words nor specific combinations of functional words were repeated. To maintain only two levels of prominence (stress/no stress), we avoided instances where unstressed words might become stressed and vice versa. This included scenarios involving focus, nominalization, citation, parentheticals, exclamatory and interrogative sentences (Face, 2003; Hualde, 2009). Additionally, compound words with two lexical stresses were also avoided.

### 2.1.3. Procedure

Participants were divided into 12 dyads: four female-female (FF), four female-male (FM), and four male-male (MM). Each dyad was tested separately in a quiet room where participants faced each other, with a 13-inch PC screen in front of each. The first member of the dyad was asked to read aloud the sentence appearing on their screen exactly as written, with a declarative intonation and neutral focus (this aspect of the procedure was implemented considering the possible role of pitch accents combined with duration as an additional acoustic correlate of speech rhythm). The second member of the dyad was instructed to listen carefully to their interlocutor (at this point, a drawing of an ear appeared on the screen) and then repeat the sentence. All participants were instructed to avoid speech overlaps and repairs.

To assess lexical accommodation between dyad members, three category-words, *poema* (poem), *tele* (short for television), and *carta* (letter), were consistently presented at the top of each participant's screen. These words appeared above the sentence to be pronounced or the drawing of an ear, indicating to listen and repeat. After each sentence was uttered by one participant and echoed by the other, both participants were instructed to categorize the sentence as related to a letter, a poem, or a TV show. They did so by verbally selecting one of the three category-words. For further details on the rationale behind this task, refer to Section 2.3.2., *Lexical repetitions*.

All four blocks of stimuli (RF, IF, RG, IG) were read/repeated by all three subgroups of dyads (FF, FM, MM). The order of reading/repetition of the blocks of stimuli by each dyad was balanced, and the order of reading/repetition of the sentences within each block was randomized. Presentation of the four blocks of stimuli was preceded by a familiarization block, consisting of six sentences that did not belong to any of the experimental blocks and lacked any particular rhythmic pattern.

At the beginning of each session, written informed consent was obtained from every participant. Participants were informed that they would take part in a study of conversations in Spanish, but no further information about the aims of the study or specific characteristics of the stimuli was provided until the end of the procedure. Sessions were recorded with an iPad Air 2 (running on iOS 11.4), equipped with a dual-microphone system that registered speakers' voices from different directions. Audio files were recorded at 44,100 Hz, 24 bits, stereo, using the software Voice Record Pro 3.3.6.

To determine whether participants noticed any differences between regular and irregular rhythmic sentences, as well as between sentences arranged in feet or in groups, participants completed a brief survey at the end of each session. The survey inquired whether they perceived any particular distinctions between the blocks of stimuli they had just read/repeated, aside from considering the meaning and order of presentation.

### 2.2. Experiment 2: Perceptual task

#### 2.2.1. Participants (evaluators)

Twenty-four native Spanish speakers—12 females and 12 males—ranging from 17 to 28 years of age participated in the study. To be eligible for the experiment, participants had to meet three conditions: (1) absence of hearing disorders; (2) no acquisition of a second language during childhood; and (3) no formal musical training. All participants were university students, and they received extra course credit as compensation for their participation. None of them had participated in Experiment 1.

#### 2.2.2. Stimuli

Audio files corresponding to pairs of sentences presented in the first, second, eighth, ninth, fifteenth, and sixteenth positions were extracted from each block of stimuli (RF, IF, RG, IG) in each of the 12 dyads' recordings from Experiment 1. This process allowed us to acquire tokens from the beginning, middle, and final parts of each dyad's interaction across all conditions, totaling 24 tokens per dyad and 288 tokens in total. Each token consisted of the same sentence produced by both participants in the dyad. In half of the tokens, participant A read the sentence and participant B repeated it; in the other half, the roles were reversed.

Using Audacity 2.0.2 for Mac, the amplitude of each token's audio signal was individually normalized, with the peak set at -1.0 dB. This process ensured that all stimuli were presented at a similar volume, while preserving the signal-to-noise ratio and internal dynamics. Additionally, half a second of total silence was inserted between each pair of sentences, replacing the original interval time between them. Another half-second of silence was added at the beginning and end of each token.

#### 2.2.3. Procedure

Each of the 24 evaluators rated a total of 96 sentence pairs, corresponding to one subgroup of dyads (FF, FM, or MM). The order in which dyads were assessed by each evaluator was counterbalanced. Additionally, both the order of block presentation (RF, IF, RG, IG) for each dyad and the order of tokens within each block were randomized.

The stimuli were presented through a pair of Sony MDR-ZX110 headphones connected to a MacBook Pro Mid 2012. Each evaluator adjusted the volume according to their preference. Evaluators were instructed to rate the rhythmic similarity between the two sentences within each token on a five-point Likert scale, with 1 indicating "not similar at all" and 5 indicating "extremely similar." They were informed that the second sentence was a repetition of the first one, and a brief explanation of speech rhythm was provided. Evaluators were specifically instructed to focus on the rhythmic aspect of the tokens, disregarding the tone of voice, volume, and speech rate of both members of the dyad. To ensure clarity, a familiarization block of six tokens was presented. These tokens were extracted from two previous pilot studies and included pairs of sentences with high, medium, and low scores in terms of their acoustic rhythmic resemblance (refer to the data analysis section). The familiarization block was scored by each evaluator under the guidance of the experimenter.

To assess whether evaluators noted any differences between regular and irregular rhythmic sentences, as well as between sentences arranged in feet or in groups, a brief survey was administered at the end of each session. Evaluators were asked to indicate if they had perceived any particular distinctions between the blocks of stimuli they had just rated, disregarding considerations related to meaning and order of presentation.

### 2.3. Data analysis

#### 2.3.1. Preparation of participants' recordings

First, the recordings were segmented into pairs of sentences and pairs of category-words. A text grid for each pair of sentences (including the silence between them) was then created using Praat 6.0.20 (Boersma & Weenink, 2017). Subsequently, text grids and corresponding audio files were aligned using the software SPPAS, version 1.7.6 (Bigi, 2015). The correction and adjustment of aligned text grids, as well as all subsequent acoustic analyses, were manually conducted by the experimenter using Praat 6.0.20.

A total of 1536 sentences were analyzed, and the following six acoustic-prosodic features were determined: (1) Total length of each sentence (seconds), (2) Length of each one of the three feet or groups within each sentence (seconds), (3) Time elapsed between the end of participant A's sentence rendition and the beginning of participant B's sentence repetition within each dyad (seconds), (4) Average pitch of each sentence (hertz), (5) Minimum pitch of each sentence (hertz), and (6) Maximum pitch of each sentence (hertz).

### 2.3.2. Determination of dependent variables

Using the aforementioned six acoustic-prosodic features, and the count of the 1536 category-words, ten dependent variables were determined, as follows:

*Interval time (IT)*: Time elapsed between the end of participant A's sentence rendition and the beginning of participant B's sentence repetition (seconds).

*F0 mean (F0M)*: Average pitch of each sentence (hertz).

*F0 range (F0R)*: Pitch range of each sentence (hertz). It was obtained by subtracting the pitch minimum from the pitch maximum in each sentence. In order to account for cross-sex comparisons, F0R was also converted to semitones applying the following formula *12 * log2 (pitch maximum / pitch minimum)*.

*Speech rate (SR):* Syllables per second, obtained by dividing the number of syllables in each sentence by its total duration. It is important to note that speech tempo can be measured in different ways, including speech rate—commonly calculated in syllables per second or words per minute—and articulation rate (Winter & Grawunder, 2012; Schultz et al., 2015). While both aim to capture aspects of temporal structure, they differ in how they treat pauses and silences. Speech rate calculations, such as the one used in this study, include these elements, which are crucial for capturing prosodic phrasing and the natural pacing of speech. In contrast, articulation rate excludes pauses and silences, thereby omitting relevant temporal information (Schultz et al., 2015).

*Rhythmic distance (RD)*: While several methods exist for measuring prosodic accommodation, there is no standardized approach (Thomason et al., 2013). For our purposes, following Späth et al.'s (2016) procedure, a rhythmic distance score was employed to determine the degree of rhythmic resemblance between each reading sentence and its repetition. This was done comparing the Euclidean distances of the relative duration of metrical units within each sentence (in our case, feet or groups). This means that for each foot/group, from each sentence, for each participant, there is a single distance measure that aids to determine the total score of each pair of sentences. Each foot or group was divided by the total length of its corresponding sentence before computing the score, normalizing each pair of sentences for speech rate. Then, the following formula was applied:

$$\sqrt{(a^1 - b^1)^\wedge 2 + (a^2 - b^2)^\wedge 2 + (a^3 - b^3)^\wedge 2}$$

Where:

$a^1, a^2, a^3$ = relative durations of feet/groups in the sentence read by the first member of the dyad (reading).

$b^1, b^2, b^3$ = relative durations of feet/groups in the sentence repeated by the second member of the dyad (repetition).

A resulting score value of 0 indicates exact equivalency between participants' metrical timing patterns (rhythms), independent of absolute speaking rate. The farther the resulting score values are from 0, the more the metrical timing patterns differ from each other. As Euclidean distances used in this way lack obvious bound values for the maximum distance but have some maximum possible discrepancy value that remains unknown until specifically computed (Barrett, 2005), convergence is quantified as a decrease in Euclidean distance, indicating an increase in similarity (Gessinger et al., 2018; Späth et al., 2016).

*F0 mean distance (F0MD)*: Considering that Euclidean distances allow the comparison of any two vectors taken across the same variable (Barrett, 2005), and have been used to establish the degree of distance/similitude of tokens for several linguistic features, such as the height and slope of pitch accents (Gessinger et al., 2018), words within a sentence (Ferrer, 2004), vowels' formant frequencies (Babel, 2010; Pardo et al., 2010), and mean pitch, mean intensity, and duration of speech preceding backchannels (Levitan et al., 2011), for our study, we propose an F0 mean distance score to assess the degree of similarity between the average F0 values of a reading sentence and its repetition. In this case, as with the rest of the distance measures, a decrease in Euclidean distance is considered an increase in similarity between the measured tokens.

However, raw Euclidean distances are sensitive to the scaling of each constituent variable, which may lead to misinterpretations when comparing variables with different score ranges, as is the case with acoustic-prosodic features of women and men (Barrett, 2005). To tackle this challenge, we individually scaled the F0 mean distance score for each of the twelve dyads using the *scale()* function in the R software.

*F0 range distance (F0RD)*: Employing a similar rationale as the F0 mean distance score, we utilized an F0 range distance score to assess the degree of distance/similarity between the F0 range values of a read sentence and its repetition. The resulting values were subsequently z-scored by dyad.

*Speech rate distance (SRD)*: We utilized a speech rate distance score based on Euclidean distances to assess the degree of distance/similarity between the speech rate values of a read sentence and its repetition. Considering that the sentences being compared are identical in terms of syllables and words, we assume that the distance between their total durations corresponds to the distance between the speech rates of the dyad members. This approach allowed us to avoid the double transformation of data from total length to syllables per second to distance score. Resulting values were subsequently z-scored by dyad.

*Lexical repetitions (LR)*: Given that the recurrent use of the same term by different individuals during a conversation to refer to an object or situation is considered an instance of lexical convergence (e.g., Brennan & Clark, 1996), and considering the potential for enormous variability in people's lexical choices, making it unlikely that two individuals use exactly the same term during a conversation to refer to a specific action or subject (Brennan, 1996), we decided to establish a small and explicit list of lexical choices to assess whether any of the conditions would lead to a greater amount of lexical repetitions between the members of each dyad. The LR variable indicates the number of times both participants classified a sentence with the same category-word: (a) *poema* (poem), (b) *tele* (short for television), or (c) *carta* (letter). In this case, the aim is not to find a “correct” category for each sentence, but rather to establish if there is a tendency to repeat the choice of the other member of the dyad (lexical convergence).

*Perceptual rating (PR)*: This variable corresponds to the rhythmic similarity ratings obtained through the five-point Likert scale used in Experiment 2, with 1 indicating "not similar at all" and 5 indicating "extremely similar."

### 2.3.3. Statistical analysis

All statistical analyses were conducted using R (R Core Team, 2017) with the packages lmerTest (Kuznetsova et al., 2017), influence.ME (Nieuwenhuis et al., 2012), and MASS (Venables & Ripley, 2002). Linear mixed-effects models were employed for the analyses, except for the lexical repetition variable (LR), which will be discussed later. P-values and Satterthwaite approximations for degrees of freedom (rounded to integers) are provided for each model. For the validity of linear mixed-effects models in Likert-scale data analysis, refer to Kizach (2014) and Norman (2010). A design-driven model selection was carried out to determine the fixed and random effects of the models (see Park et al., 2020). The inclusion of subject-specific (dyad or participant) or item-specific slopes was determined on a case-by-case basis, identifying the largest model that converged and comparing it to the intercept-only model, as described by Barr (2013).

Five variables were included as fixed-effect factors in the analysis: (1) rhythmic regularity, coded as STRCT, with two levels: regular/irregular; (2) type of phrasing, coded as UNIT, with two levels: groups/feet; (3) type of dyad, coded as SEX, with three levels: FF (female-female), FM (female-male), and MM (male-male); (4) half of the test (i.e., first or last half of the 16 sentences comprising each block of stimuli), coded as HALF, with two levels: first/last, and (5) mode of rendition, coded as MODE, with two levels: reading/repetition.

To address the independence of observations, two variables were employed as crossed random effects (Baayen, 2008). For models with dependent variables obtained through paired measures (interval time and distance scores), as well as the perceptual rating model, the crossed random effects included: (1) the 64 sentences used as stimuli, coded as ITEM, with one level for each sentence; and (2) the 12 dyads, coded as DYAD, with one level for each dyad. For models with a dependent variable obtained through individual measures (F0 range, F0 mean, and speech rate), the crossed random effects comprised: (1) the 64 sentences used as stimuli, coded as ITEM, the same as in the other models; and (2) the 24 individual participants, coded as PTCP, with one level for each.

Since the raw data did not reveal substantial differences in lexical repetitions concerning STRCT or UNIT, a binomial logistic regression was employed to assess the significance of observed variations between types of dyads (SEX) (note that the data for the three dyad types, FF, FM, and MM, consist of independent observations). In this context, P-values are determined through likelihood ratio chi-square tests.

### 2.4. Hypothesized results

Considering previous findings showing that listeners extract rhythmic beats from their interlocutors' utterances during conversation—beats that, in turn, influence their own productions (Schultz et al., 2015); that listeners process information more quickly within syllables expected to bear stress based on prior prosodic context (Brown et al., 2015); that repeated exposure to speech reduces processing time and neural activation, while multiple repetitions enhance memory and learning (Falk et al., 2014); and that speech rhythm resemblance is greater in metrically regular sentences than in irregular ones (Späth et al., 2016)—we expect the following effects to emerge in our study.

During interactions involving regular rhythmic sentences, the time between hearing and repeating utterances should be shorter (H1); distance measures between speakers should be smaller (H2); and lexical repetitions should be more frequent (H3)—compared to interactions involving irregular rhythmic sentences. Furthermore, considering that in South American Spanish, accentual groups tend to be more uniform in duration than syllables or accentual feet (Toledo, 1988), these effects should be stronger in interactions involving sentences arranged in accentual groups than in those arranged in accentual feet (H4).

In addition, the study examines variation in F0 range and mean, speech rate, and mode of rendition (reading vs. repetition) across rhythmic (STRCT) and phrasing (UNIT) conditions, providing novel insights into how these variables may relate to interlocutor accommodation.

## 3. Results

Next, each dependent variable is discussed individually. Table 1 presents an overview of the results.

| | | Rhythmic regularity (STRCT) | | Phonological phrasing (UNIT) | | Half of the test (HALF) | | Mode of rendition (MODE) | | Type of Dyad (SEX) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Regular | Irregular | Groups | Feet | First | Last | Read. | Repeat. | Female only | Mixed | Male only |
| RD | Rhythmic dist. | 0.053 | 0.065 | 0.060 | 0.058 | 0.058 | 0.060 | — | — | 0.053 | 0.064 | 0.060 |
| IT | Interval time (s) | 0.613 | 0.664 | 0.660 | 0.617 | 0.642 | 0.635 | — | — | 0.600 | 0.575 | 0.740 |
| F0R | F0 range (Hz) | 78.96 | 88.30 | 80.36 | 86.91 | 84.94 | 82.33 | 91.43 | 75.84 | 103.13 | 86.42 | 61.35 |
| F0RD | F0 range dist. | 1.076 | 1.201 | 1.076 | 1.201 | 1.123 | 1.153 | — | — | 0.967 | 1.316 | 1.132 |
| F0M | F0 mean (Hz) | 181.56 | 183.62 | 180.77 | 184.42 | 182.57 | 182.61 | 185.88 | 179.30 | 227.64 | 189.78 | 130.35 |
| F0MD | F0 mean dist. | 1.561 | 1.629 | 1.583 | 1.607 | 1.587 | 1.603 | — | — | 1.371 | 1.943 | 1.472 |
| SR | Speech rate (σ/s) | 5.42 | 5.28 | 5.45 | 5.25 | 5.40 | 5.30 | 5.17 | 5.53 | 4.93 | 5.39 | 5.73 |
| SRD | Speech rate dist. | 0.94 | 1.00 | 0.98 | 0.96 | 0.97 | 0.97 | — | — | 0.97 | 1.04 | 0.90 |
| LR | Lexical rep. (n) | 158 | 163 | 164 | 157 | 150 | 171 | — | — | 116 | 91 | 114 |
| PR | Percept. rating (Likert) | 3.22 | 3.20 | 3.22 | 3.20 | 3.29 | 3.14 | — | — | 3.07 | 3.56 | 3.01 |

**Table 1.** General results. All distance scores are based on Euclidean distances, where lower scores indicate a higher degree of similarity.

### 3.1. Experiment 1: Acoustic task

### 3.1.1. Rhythmic distance (RD)

There was a significant effect of the factor STRCT ($F(1,60) = 22.59$; $p < 0.001$), but not of UNIT ($F(1,60) = 0.62$; $p = 0.433$), SEX ($F(2,9) = 1.61$; $p = 0.253$), or HALF ($F(1,749) = 0.93$; $p = 0.336$). Additionally, there was a significant interaction effect for STRCT × UNIT ($F(1,60) = 4.50$; $p < 0.05$). After a square root data transformation applied to resolve a deviation from homoscedasticity a significant interaction of SEX × HALF ($F(2,750) = 3.54$; $p = 0.029$) was lost, and was replaced by ($F(2,750) = 2.67$; $p = 0.069$). Hence, rhythmic distance was closer in metrically regular sentences compared to irregular sentences, and this effect was greater in sentences arranged by groups than in sentences arranged by feet (see Figure 1).

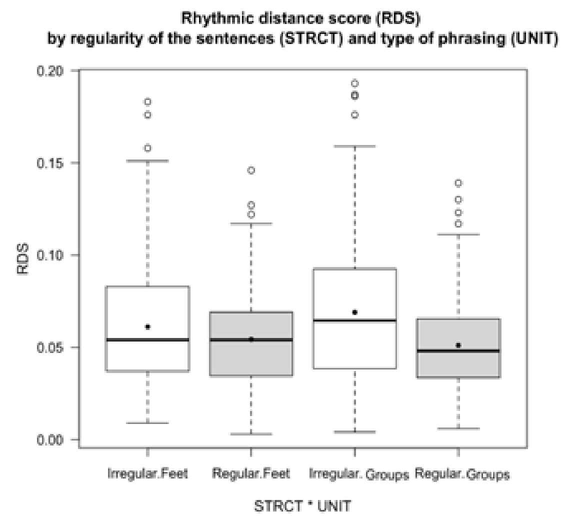


**Figure 1.** Rhythmic distance score (RDS) by regularity of the sentences (STRCT) and type of phrasing (UNIT).

### 3.1.2. Interval time (IT)

There were no significant effects of any of the factors: STRCT ($F(1,30) = 2.09$; $p = 0.159$), UNIT ($F(1,23) = 2.57$; $p = 0.123$), SEX ($F(2,15) = 0.86$; $p = 0.442$), or HALF ($F(1,716) = 0.08$; $p = 0.779$). Also, no significant interactions between predictors were found.

### 3.1.3. F0 range (F0R)

There were significant effects of the factors STRCT ($F(1,48) = 10.45$; $p < 0.01$), UNIT ($F(1,57) = 7.81$; $p < 0.01$), SEX ($F(2,21) = 7.87$; $p < 0.01$), HALF ($F(1,1421) = 11.32$; $p < 0.001$), and MODE ($F(1,1376) = 288.55$; $p < 0.001$). Additionally, a significant interaction effect was found for the factors HALF × SEX ($F(2,1424) = 3.04$; $p < 0.05$). Consequently, participants' F0 ranges were narrower in metrically regular sentences compared to irregular ones, in sentences arranged by group compared to sentences arranged by feet, and during repetition compared to reading. Also, F0 ranges were narrower in male-only dyads compared to female-only.

Additionally, participants' F0 ranges were narrower during the last half of the interaction, especially in female-only dyads, and to a lesser extent in mixed dyads. Conversely, participants' F0 ranges were slightly wider during the last half of the interaction in male-only dyads (see Figure 2). In this respect, when the model was calculated without female-only dyads, the significance of SEX disappeared ($F(1,14) = 3.39$; $p = 0.087$), as well as the interaction effect for SEX × HALF ($F(1,955) = 1.96$; $p = 0.162$) (All the other interactions remaining). This fact indicates that the SEX × HALF interaction, as well as the main difference in SEX regarding F0 range, must only be considered for female-female dyads. Furthermore, when measured in semitones, there was a significant difference between the F0R of women (M = 7.67, SD = 2.18) and men (M = 8.13, SD = 2.35), t(1525.2) = -4.02, $p = 6.024e-05$.

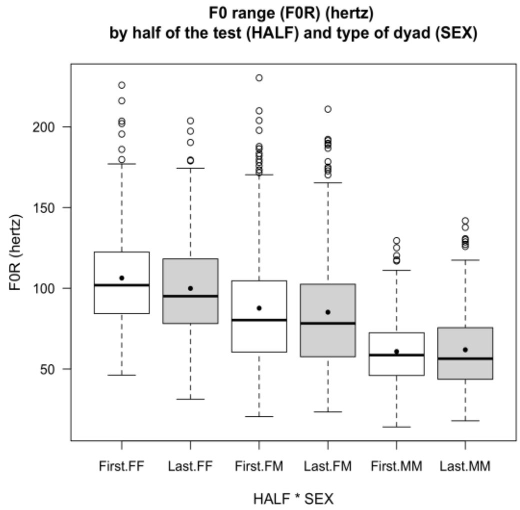


**Figure 2.** F0 range (F0R) (hertz) by half of the test (HALF) and type of dyad (SEX) (FF: Female-Female; FM: Female-Male; MM: Male-Male).

### 3.1.4. F0 range distance (F0RD)

There was a significant effect of the factor UNIT ($F(1,61) = 4.98$; $p < 0.05$), but not of STRCT ($F(1,61) = 1.79$; $p = 0.187$), SEX ($F(2,9) = 1.76$; $p = 0.226$), or HALF ($F(1,747) = 0.01$; $p = 0.939$). Additionally, there was a significant interaction effect for STRCT × SEX ($F(2,690) = 5.56$; $p < 0.01$). Hence, the distance between interlocutors' F0 ranges was closer in sentences arranged by groups than in sentences arranged by feet. Additionally, F0 range distances were closer in metrically regular sentences compared to irregular ones, especially in mixed dyads, and to a lesser extent in female-only dyads. Conversely, metrically irregular sentences were slightly closer in terms of F0 range distances with respect to regular sentences in male-only dyads (see Figure 3). When the model was calculated without mixed dyads, the interaction effect for STRCT × SEX disappeared ($F(1,439) = 0.57$; $p = 0.45$) (With the UNIT interaction remaining). This fact indicates that the STRCT × SEX effect must only be considered for female-male dyads.

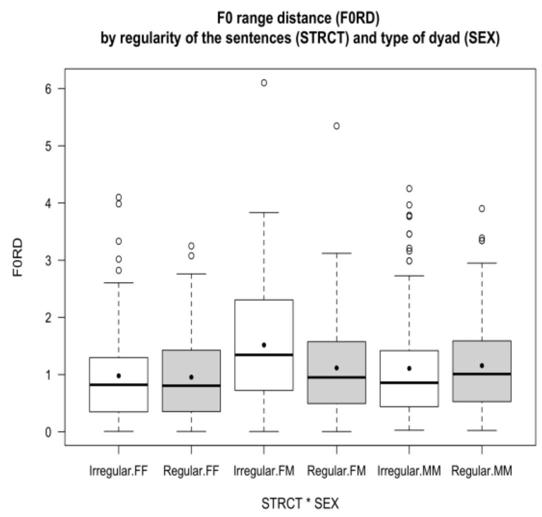


**Figure 3.** F0 range distance (F0RD) by regularity of the sentences (STRCT) and type of dyad (SEX) (FF: Female-Female; FM: Female-Male; MM: Male-Male).

### 3.1.5. F0 mean (F0M)

There were significant effects of the factors STRCT ($F(1,50) = 7.45$; $p < 0.01$), UNIT ($F(1,47) = 11.13$; $p < 0.01$), SEX ($F(2,21) = 13.84$; $p < 0.001$), and MODE ($F(1,1378) = 478.48$; $p < 0.001$), but not of HALF ($F(1,1418) = 0.07$; $p = 0.794$). Additionally, a significant interaction effect was observed for MODE × UNIT ($F(1,1378) = 10.99$; $p < 0.001$). Consequently, participants' F0 means were lower in metrically regular sentences compared to irregular ones, and in sentences arranged by group compared to those arranged by feet. Moreover, participants' F0 means were lower in male-only dyads compared to female-only dyads, and F0 means were lower during repetition than during reading, with a greater effect observed in sentences arranged by feet compared to those arranged by groups.

### 3.1.6. F0 mean distance (F0MD)

There was a significant effect of the factor SEX ($F(2,13) = 7.68$; $p < 0.01$), but not of STRCT ($F(1,12) = 0.13$; $p = 0.726$), UNIT ($F(1,15) = 0.00$; $p = 0.953$), or HALF ($F(1,725) = 0.01$; $p = 0.911$). No significant interactions between predictors were found. Hence, the distance between interlocutors' F0 means was closer in female-only and male-only dyads compared to mixed dyads. This is an expected result, as mixed dyads' scores are calculated by comparing F0 means of women with F0 means of men, which are "far" from each other in terms of Euclidean distances and "further away" compared to same-sex dyads. Moreover, when the model was calculated without mixed dyads, the significant effect of the factor SEX was lost ($F(1,9) = 0.58$; $p = 0.465$), indicating no difference between same-sex dyads in terms of F0 mean distance.

### 3.1.7. Speech rate (SR)

There were significant effects of the factors SEX ($F(2,21) = 4.23$; $p < 0.05$), HALF ($F(1,1388) = 11.11$; $p < 0.001$), and MODE ($F(1,1376) = 251.55$; $p < 0.001$), but not of STRCT ($F(1,72) = 1.65$; $p = 0.203$) or UNIT ($F(1,70) = 3.33$; $p = 0.072$). Additionally, a significant interaction effect for HALF × SEX was observed ($F(2,1389) = 3.93$; $p < 0.05$). Hence, participants spoke faster during sentence repetition than during reading, and in the male-only dyads compared to the female-only dyads. Moreover, speech rate was higher during the first half of the interaction, particularly in male-only dyads and, to a lesser extent, in mixed dyads. Conversely, speech rate was slightly lower during the first half of the interaction in female-only dyads (see Figure 4).

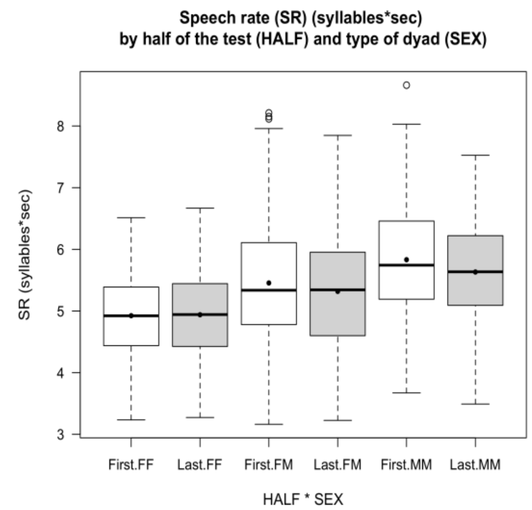


**Figure 4.** Speech rate (SR) (syllables/sec) by half of the test (HALF) and type of dyad (SEX) (FF: Female–Female; FM: Female–Male; MM: Male–Male).

### 3.1.8. Speech rate distance (SRD)

No significant effects were observed for any of the factors: STRCT ($F(1,61) = 0.50$; $p = 0.484$), UNIT ($F(1,61) = 0.12$; $p = 0.730$), SEX ($F(2,9) = 0.50$; $p = 0.620$), or HALF ($F(1,749) = 0.01$; $p = 0.919$). Additionally, no significant interactions between predictors were found.

### 3.1.9. Lexical repetitions (LR)

The factor SEX showed statistical significance ($\chi^2 = 6.2778$, $df = 2$, $p < 0.05$), suggesting a higher occurrence of lexical repetitions between participants in same-sex dyads compared to participants in mixed-sex dyads.

### 3.1.10. Participants' survey

Here, we present key remarks made by participants, each followed by a numerical indication of how many participants shared a particular observation. Participants referred to the stimulus blocks as "the first one," "the one in the middle," or similar terms, as they were not aware of the actual names of the blocks. Some common observations include the difficulty in uttering (1), reading (8), or repeating (3) articles such as *las* and pronouns like *les* and *nos*. Notably, these articles and pronouns were present in all stimulus blocks, but certain combinations within irregular groups and regular feet seemed to pose challenges for the participants. Other observations highlighted the difficulty in reading regular feet (3), irregular groups (2), or irregular feet (2). Surprisingly, irregular groups were considered easier to utter (2) or read (1).

## 3.2. Experiment 2: Perceptual task

### 3.2.1. Perceptual rating (PR)

There were significant effects observed for the factors SEX ($F(2,10) = 11.71$; $p < 0.01$) and HALF ($F(1,301) = 5.86$; $p < 0.05$), while no significant effects were found for STRCT ($F(1,16) = 0.47$; $p = 0.503$) or UNIT ($F(1,11) = 0.11$; $p = 0.751$). Additionally, a significant interaction effect was identified for UNIT × SEX ($F(2,11) = 4.58$; $p < 0.05$).

Hence, sentences of mixed dyads were rated more similar to each other with respect to female-only and male-only dyads, and sentences uttered during the first half of the interaction were rated more similar to each other with respect to the last half. Additionally, in mixed and male-only dyads sentences were rated more similar to each other in terms of rhythm when they were arranged by feet, while in female-only dyads they were rated more rhythmically similar when arranged by groups. The results indicate that sentences in mixed dyads were rated more similar to each other compared to female-only and male-only dyads. Moreover, sentences uttered during the first half of the interaction were rated more similar to each other than those in the last half. Notably, in mixed and male-only dyads, participants tended to rate sentences as more rhythmically similar when arranged by feet, while in female-only dyads, a higher perceived rhythmic similarity was associated with sentences arranged by groups.

### 3.2.2. Evaluators' survey

Here, we present key remarks made by evaluators, each followed by a numerical indication of how many evaluators shared a particular observation. It is important to note that evaluators participating in the perceptual task were not aware that the first sentence of each pair was read. Observations included instances where some individuals omitted the final “s” in certain words (5), added a final “s” to some words (1), or lengthened the final “s” in specific words (1). Additionally, evaluators noted variations in the pace of speech (6), with the understanding that this refers to differences between individuals rather than differences between types of sentences. Other observations included changes in words during repetitions (4), mistakes made by some participants (3), and the perception that some utterances were more challenging to articulate than others (2).

## 4. Discussion

While the influence of speech rhythm on linguistic interactions is well-established, the underlying mechanisms remain elusive, particularly in Spanish-speaking contexts. In this study, we examined the impact of speech rhythm on linguistic accommodation during dyadic interactions, examining both its acoustic and perceptual dimensions.

Our results replicated previous studies, showing that men exhibit faster speech rates than women (Cohen et al., 2017; Weirich & Simpson, 2014). However, contrary to previous findings suggesting that spontaneous speaking tends to be slower than reading in both women and men (Lass & Sandusky, 1971; Snidecor, 1943), our findings revealed that participants spoke faster during sentence repetition than during reading. Crucially, all sentences in our study, whether spoken or read, were of equal length and lacked punctuation or other indicators of pauses. Despite this uniformity, participants introduced "irregular" pauses within some sentences, as indicated by acoustic analyses and participant reports in the survey following the perceptual task. From a cognitive perspective, it is essential to recognize that repeating and reading are not totally equivalent tasks, since repetitions, benefiting from the recent mental activation of the same sentence, tend to be faster than readings.

We also found that female-only dyads exhibited narrower F0 ranges and higher speech rates during the interaction's latter half compared to the initial phase. This is one of the few time-related accommodation effects identified in this study and might be attributed to task-induced fatigue. In this context, the speech rate might have increased towards the end of the task to expedite completion, and the F0 range may have become flatter, as commonly observed in the speech of fatigued individuals. Another possibility is that the observed increase in speech rate during the second half of the interaction was a consequence of accommodation taking place during the first half. This explanation aligns with the finding that sentences uttered early in the interaction were perceptually rated as more similar to each other compared to the later sentences. However, no effects of HALF (the first or last half of the test) were observed in any of the distance measures. Additionally, male-only, and mixed dyads exhibited the opposite pattern, tending to speak slower towards the end of the task.

We also observed a higher degree of lexical convergence, measured as word repetitions, between participants in same-sex dyads compared to mixed dyads. This influence of the interlocutors' gender on dyadic accommodation is partially consistent with Street's (1984) study, where male-only dyads tended to converge in turn duration, while female-male dyads tended to diverge. Simultaneously, in our perceptual task experiment, sentences from mixed dyads were rated as more similar to each other compared to sentences from female-only and male-only dyads. Besides, similar to some other studies (e.g., Kawasaki et al., 2013; Thomson et al., 2001), no further significant differences between men and women in terms of accommodation were observed.

A crucial result of this research is the observed increase in rhythmic resemblance—reflected in smaller Euclidean distances—among Spanish speakers facilitated by sentence-level temporal regularity (supporting H2; see Section 2.4). This finding aligns with the results presented by Späth et al. (2016) for German healthy speakers and patients with Parkinson's disease. Also, the heightened effect of rhythmic distance being closer in metrically regular sentences—especially in those arranged by accentual groups—supports H4 and is consistent with previous reports indicating greater temporal regularity in accentual groups compared to accentual feet in Spanish (Almeida, 1997; Mora et al., 1999; Toledo, 1988).

A potential explanation for accentual groups exhibiting closer rhythmic distances lies in the concept of *sirrema* proposed by Quilis (1993). According to this concept, a sirrema is a group of two or more words forming a grammatical, melodic, and semantic unit. Various combinations, such as article + noun, noun + adjective, and conjunctions + the connected element, can constitute a sirrema. Importantly, there is typically no pause between the words within a sirrema during utterance under normal circumstances (Spang, 1983).

In this scenario, structural relations between the components of a sentence determine to a certain extent the way in which such sentence is phrased. Unlike accentual feet, which depend mostly on metric relations at the phonological level, accentual groups also rely on syntactic connections between words. Therefore, it is possible that the typical phrasing of the accentual group, deep-rooted in Spanish speakers, might have impeded the "correct" phrasing of the Regular feet block of stimuli —[100] [100] [100]—, resulting in an accentual non-regular phrasing and, consequently, less rhythmic resemblance.

Regarding accommodation theories, the significant impact of metrical regularity within accentual groups, which depend on relations across different linguistic levels, and the fact that post-experiment surveys suggest participants were not consciously aware of the phrasing and regularity dichotomy, both align with the idea that rhythmic accommodation operates through a multilevel non-conscious mechanism, such as the one described by the interactive alignment model (IAM) (Garrod & Pickering, 2004; Xu & Reitter, 2016). In this context, the stronger impact of regularity on increasing rhythmic resemblance within accentual groups may have contributed to the heightened similarity observed in the F0 ranges between speakers within these groups. Similarly, the same rationale could be applied to the regular/irregular distinction, although in this case, only mixed dyads showed closer pitch range distances in metrically regular sentences.

Apart from the results just mentioned, no other hypotheses regarding the effects generated by conversational interactions involving regular rhythmic sentences and sentences arranged in accentual groups were confirmed, including the absence of significant differences in the number of lexical repetitions between rhythmic conditions (offering no support for H3). This lack of significant results could be due to at least two reasons: there were in fact no differences between the experimental conditions with respect to such variables, or, more likely, the methodological approach employed in this research was unable to determine further differences between conditions.

Similarly, no differences were found in response times between hearing and repeating sentences cross conditions, suggesting that processing different rhythmic regularities and phonological phrasing may rely on relatively automatic mechanisms, involving a consistent cognitive load across conditions (no support for H1). However, this finding diverges from previous research reporting faster processing of metrically regular utterances and quicker access to syllables expected to bear stress based on rhythmic patterns (Brown et al., 2015; Mooney & Sullivan, 2015).

In this respect, we acknowledge that reading and pronunciation difficulties reported by some participants may have influenced overall response times, despite our statistical analysis precautions. To address this, involving model talkers in future research could help maintain the intended "pure" rhythmic patterns. Alternatively, testing participants with a computer-generated perfectly regular rhythm might offer a controlled environment for studying pure rhythmic patterns.

It is noteworthy that no convergence effects were observed during the experiments. Instead, sentences uttered during the first half of the interaction were perceptually rated as more similar to each other compared to sentences uttered during the last half. It is important to highlight that the measurement of accommodation in this study was based on the concept of similarity rather than convergence. There are alternative approaches that focus on the temporal development of accommodation, and some of these may be considered for future research (see De Looze et al., 2014, for an overview). Despite this, the results concerning the effect of HALF (half of the test) and visual inspections of the similarity levels between speakers during each experimental block did not indicate an increase in any of the tested variables.

Looking from another perspective, the limited definition of rhythm used in this study (specifically, rhythm as a temporal iteration of strong and weak values of stressed and unstressed syllables) may be considered a limitation of our approach. For instance, such a definition might not be suitable for tone or pitch-accent languages. Future research would benefit from analyses based on broader definitions of speech rhythm. Additionally, exploring different types of feet and group arrangements, including materials where feet are completely enclosed within words, could provide valuable insights.

Regarding the coherence between acoustic and perceptual measurements, our experiments revealed contrasting patterns. Sentences were perceived as more similar when arranged by feet and spoken by mixed or men-only dyads. Conversely, acoustic analyses indicated greater rhythmic similarity in sentences arranged by groups and uttered by women-only dyads. These conflicting results contribute to the existing inconsistencies between acoustic and perceptual data in studies of linguistic accommodation, which were mentioned earlier. Likewise, they complicate the interpretation of H2, highlighting the need for caution when inferring perceptual similarity solely from acoustic distance.

In our study, although participants were explicitly instructed to focus on rhythmic differences, their judgments likely reflected a broader range of acoustic characteristics beyond the intended rhythmicity. It is also possible that they relied on dimensions that varied systematically across utterances, such as pitch, articulation, or intensity. Moreover, perceived rhythmic salience, rather than actual rhythmic similarity, may have driven their ratings. These considerations suggest that a different experimental design would be required to more confidently determine whether temporal regularity directly influences perceived rhythmic similarity.

Finally, our data revealed that pitch ranges were narrower, and the average pitch was lower in metrically regular sentences compared to irregular ones, as well as in sentences arranged by groups compared to those arranged by feet. Although no specific hypotheses were formulated regarding these acoustic-prosodic features, these results may reflect the influence of temporal regularity and phonological phrasing on vocal dynamics. One plausible explanation is that the greater temporal regularity of sentences arranged by groups induced a habituation effect, making the speech less dynamic and resulting in a flatter pitch range. This flatter pitch range, in turn, may account for the lower average pitch.

## 5. Conclusion

Our findings indicate that temporally regular iterations of stressed and unstressed syllables, organized into accentual groups, promote greater speaker resemblance in rhythm and F0 range compared to irregular iterations structured by accentual feet. Furthermore, the results show that both rhythmic regularity and phonological phrasing influence not only similarity between speakers, but also the mean and range of their fundamental frequency during conversational interactions. Regarding the psychological basis of linguistic coordination, the significant impact of metrical regularity within accentual groups—whose structure depends on relationships across multiple linguistic levels—combined with participants' lack of conscious awareness of rhythmic patterns (as indicated in post-experiment surveys), supports the notion that rhythmic accommodation operates through a multilevel, non-conscious mechanism, consistent with the interactive alignment model.

## References

Alcoba, S. (2007). Usos de *cual*, grupo acentual y unidad melódica. *Moenia*, *13,* 39–68.

Almeida, M. (1993). Alternancia temporal y ritmo en español. *Verba*, *20*, 433–443.

Almeida, M. (1997). Organización temporal del español: El principio de isocronía. *Revista de Filología Románica*, *14*(1), 29–40.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1–2), 46–63. https://doi.org/10.1159/000208930

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. https://doi.org/10.1017/s0047404510000400

Babel, M. (2011). Imitation in speech. *Acoustics Today*, *7*(4), 16–22. https://doi.org/10.1121/1.3684224

Barón, L. (2018). *Influence of sentence-level rhythmic regularity and phonological phrasing on linguistic accommodation during conversational interactions: The case of Spanish-speaking dyads* [Doctoral dissertation, Université d'Aix-Marseille]. HAL. https://theses.hal.science/tel-01993542/

Barón, L. (2023). Phonetic accommodation during conversational interactions: An overview. *Revista Guillermo de Ockham*, *21*(2), 493–517. https://doi.org/10.21500/22563202.6150

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4,* Article 348. https://doi.org/10.3389/fpsyg.2013.00328

Barrett, P. (2005). *Euclidean distance: Raw, normalized, and double-scaled coefficients* [White paper]. http://www.pbarrett.net/techpapers/euclid.pdf

Bigi, B. (2015). SPPAS - Multi-lingual approaches to the automatic annotation of speech. *The Phonetician*, *111–112*, 54–69. https://hal.science/hal-01417876v1

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.20) [Computer software]. http://www.praat.org/

Borzone, A., & Signorini, A. (1983). Segmental duration and rhythm in Spanish. *Journal of Phonetics, 11*(2), 117–128. https://doi.org/10.1016/s0095-4470(19)30810-1

Brennan, S. (1996). Lexical entrainment in spontaneous dialog. In H. Fujisaki (Ed.), *Proceedings of the International Symposium on Spoken Dialogue* (pp. 41–44). Acoustical Society of Japan.

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brown, M., Salverda, A., Dilley, L., & Tanenhaus, M. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 306–323. https://doi.org/10.1037/a0038689

Cantero, F. J. (2002). *Teoría y análisis de la entonación*. Edicions de la Universitat de Barcelona.

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio–motor training affect speech perception. *Acta Psychologica*, *155*, 43–50. https://doi.org/10.1016/j.actpsy.2014.12.002

Cohen, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *The Journal of the Acoustical Society of America*, *141*(5), 2989–2996. https://doi.org/10.1121/1.4982199

Cummins, F. (2009). Rhythm as an affordance for the entrainment of movement. *Phonetica*, *66*(1–2), 15–28. https://doi.org/10.1159/000208928

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*(1), 51–62. https://doi.org/10.1016/s0095-4470(19)30776-4

De Looze, C., & Rauzy, S. (2011). Measuring speakers' similarity in speech by means of prosodic cues: Methods and potential. In P. Cosi, & R. De Mori (Eds.), *Proceedings of Interspeech 2011* (pp. 1393–1396). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2011-457

De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, *58*, 11–34. https://doi.org/10.1016/j.specom.2013.10.002

Face, T. (2003). Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics*, *2*, 115–131. https://doi.org/10.5565/rev/catjl.46

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1491–1506. https://doi.org/10.1037/a0036858

Ferrer, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, *70* (5), article 056135. https://doi.org/10.1103/physreve.70.056135

Garrod, S., & Pickering, M. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Gessinger, I., Schweitzer, A., Andreeva, B., Raveh, E., Möbius, B., & Steiner, I. (2018). Convergence of pitch accents in a shadowing task. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Speech Prosody 2018* (pp. 225–229). International Speech Communication Association. https://doi.org/10.21437/speechprosody.2018-46

Gibbon, D. (2015). Speech rhythms: Modeling the groove. In R. Vogel & R. van de Vijver (Eds.), *Rhythm in cognition and grammar: A Germanic perspective* (pp. 108–161). De Gruyter Mouton. https://doi.org/10.1515/9783110378092.53

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics. Studies in emotion and social interaction* (pp. 1–68). Cambridge University Press. https://doi.org/10.1017/cbo9780511663673.001

Grau, A. (2013). Reconsidering syllabic minimality in Spanish truncation. *ELUA: Estudios de Lingüística Universidad de Alicante*, *27*, 121–143. https://doi.org/10.14198/elua2013.27.05

Harris, M. J. (2015). *Quantifying speech rhythms: Perception and production data in the case of Spanish, Portuguese, and English* [Doctoral dissertation, University of California, Santa Barbara]. ProQuest Dissertations Publishing. (Publication No. 3689928). http://search.proquest.com/docview/1679467612

Heath, J. (2015). Convergence through divergence: Compensatory changes in phonetic accommodation. *LSA Annual Meeting Extended Abstracts*, *6*, Article 7. https://doi.org/10.3765/exabs.v0i0.3002

Hualde, J. (2009). Unstressed words in Spanish. *Language Sciences*, *31*(2–3), 199–212. https://doi.org/10.1016/j.langsci.2008.12.003

Hualde, J., & Nadeu, M. (2014). Rhetorical stress in Spanish. In H. van der Hulst (Ed.), *Word stress: Theoretical and typological issues* (pp. 228–252). Cambridge University Press. https://doi.org/10.1017/cbo9781139600408.010

Jones, M, Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*(4), 313–319. https://doi.org/10.1111/j.0956-7976.2002.00458.x

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., & Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific Reports*, *3*, Article 1692. https://doi.org/10.1038/srep01692

Kizach, J. (2014, February). *Analyzing Likert-scale data with mixed-effects linear models: A simulation study* [Poster presentation]. Linguistic Evidence 2014, Tübingen, Germany.

Kuznetsova, A. Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lass, N., & Sandusky, J. (1971). A study of the relationship of diadochokinetig rate, speaking rate and reading rate. *Today's Speech, 19*(3), 49–54. https://doi.org/10.1080/01463377109368992

Levitan, R., Gravano, A., & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In D. Lin, Y. Matsumoto, & R. Mihalcea (Eds.), *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 113–117). Association for Computational Linguistics. https://aclanthology.org/P11-2020/

Mooney, S., & Sullivan, G. (2015). Investigating an acoustic measure of perceived isochrony in conversation: Preliminary notes on the role of rhythm in turn transitions. *University of Pennsylvania Working Papers in Linguistics, 21*(2), 128–135. https://repository.upenn.edu/handle/20.500.14332/45060

Mora, E., Villamizar, T., Blondet, M., & López, Y. (1999). Hacia una caracterización rítmica del español hablado en Venezuela. *Boletín Antropológico*, *47*, 75–87.

Nieuwenhuis, R., Grotenhuis, M., & Pelzer, B. (2012). influence.ME: Tools for detecting influential data in mixed effects models. *R Journal*, *4*(2), 38–47. https://doi.org/10.31235/osf.io/a5w4u

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Pardo, J. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. (2013). Reconciling diverse findings in studies of phonetic convergence. In *Proceedings of Meetings on Acoustics*, *19*(1), Article 060140. https://doi.org/10.1121/1.4798479

Pardo, J., Jay, I., & Krauss, R. (2010). Conversational role influences speech imitation. *Attention,*

*Perception, & Psychophysics, 72*(8), 2254–2264. https://doi.org/10.3758/app.72.8.2254
Park, J., Cardwell, R., & Yu, H. T. (2020). Specifying the random effect structure in linear mixed effect models for analyzing psycholinguistic data. *Methodology*, *16*(2), 92–111. https://doi.org/10.5964/meth.2809
Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056
Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Gredos.
R Core Team. (2017). *R: A language and environment for statistical computing* (Version 3.x) [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/
Real Academia Española. (2008). *Corpus de referencia del español actual (CREA)* [Corpus]. http://corpus.rae.es/creanet.html
Ruch, H., Zürcher, Y., & Burkart, J. (2018). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*, *93*(2), 996–1013. https://doi.org/10.1111/brv.12382
Schultz, B., O'Brien, I., Phillips, N., McFarland, D., Titone, D., & Palmer, C. (2015). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics*, *37*(5), 1201–1220. https://doi.org/10.1017/s0142716415000545
Snidecor, J. (1943). A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading. *Speech Monographs*, *10*(1), 50–56. https://doi.org/10.1080/03637754309390077
Spang, K. (1983). *Ritmo y versificación. Teoría y práctica del análisis métrico y rítmico*. Universidad de Murcia.
Späth, M., Aichert, I., Ceballos, A., Wagner, E., Miller, N., & Ziegler, W. (2016). Entraining with another person's speech rhythm: Evidence from healthy speakers and individuals with Parkinson's disease. *Clinical Linguistics and Phonetics*, *30*(1), 68–85. https://doi.org/10.3109/02699206.2015.1115129
Street, R. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research, 11*(2), 139–169. https://doi.org/10.1111/j.1468-2958.1984.tb00043.x
Thomason, J., Nguyen, H., & Litman, D. (2013). Prosodic entrainment and tutoring dialogue success. In H. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education* (pp. 750–753). Springer. https://doi.org/10.1007/978-3-642-39112-5_104
Thomson, R., Murachver, T., & Green, J. (2001). Where is the gender in gendered language? *Psychological Science, 12*(2), 171–175. https://doi.org/10.1111/1467-9280.00329
Toledo, G. (1988). *El ritmo en el español: Estudio fonético con base computacional*. Gredos.
Toledo, G. (2009). Métricas rítmicas en tres dialectos Amper-Hispanoamérica. *Ianua: Revista Philologica Romanica*, *9*, 1–21.
Toledo, G. (2010). Métricas rítmicas en microdiscursos. *Onomázein*, *21*, 71–95. https://doi.org/10.7764/onomazein.21.03
Venables, W., & Ripley, B. (2002). *Modern applied statistics with S*. Springer. https://doi.org/10.1007/978-0-387-21706-2
Weirich, M., & Simpson, A. (2014). Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics, 43*, 1–10. https://doi.org/10.1016/j.wocn.2014.01.001
Winter, B., & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics, 40*(6), 808–815. https://doi.org/10.1016/j.wocn.2012.08.006
Xu, Y., & Reitter, D. (2016). Convergence of syntactic complexity in conversation. In K. Erk, & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 443–448). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-2072

**Appendix:** Complete list of sentences used in Experiment 1

Regular groups

| | |
|---|---|
| El padre te quiere con ganas | The father likes you a lot |
| El carro los trae de vuelta | The car brings them back |
| La lluvia se siente de cobre | The rain feels like cooper |
| La madre los llama sin cargo | The mother calls them for free |
| Las noches les pasan de largo | The nights passed them by |
| Las niñas les muestran las caras | The girls show their faces to him |
| Los casos le salen por miles | The cases are numerous for him |
| Los malos lo sufren sin rabia | Bad people suffer it without rage |
| Los niños nos tienen tu gato | The kids keep your cat for us |
| Mi prima le saca las fotos | My cousin takes his pictures |
| Mi novia se lleva tu nota | My girlfriend takes your note away |
| Su perro te llena de barro | His dog gets you muddy |
| Su casa se cae por partes | His house falls apart |
| Tu lora lo dijo con gracia | Your parrot said it gracefully |
| Tu primo nos paga las cuentas | Your cousin pays us the bills |
| Tus gastos nos dejan sin plata | Your expenses leave us without money |

Irregular groups

| | |
|---|---|
| Alba me le subió de precio | Alba raised the price for me |
| Ana me les cambió la vida | Ann changed their lives for me |
| Ángel te lo pasó de lado | Angel passed it edgewise to you |
| Carlos te la mostró con fotos | Charles showed it to you with photos |
| Carmen te las perdió por miedo | Carmen lost them for fear |
| Clara te los jugó sin ganas | Clara unwillingly played them to you |
| Ella nos los negó por plata | She denied them to us for money |
| Ellas me los tendrán si quieren | They will keep them for me if they want |
| Laura nos lo pidió sin gracia | Lora gracelessly asked us for it |
| Mario se nos quedó sin casa | Mario was left homeless |
| Marta se las dará si paga | Martha will give them to her if she pays |
| Pablo se le salió con rabia | Paul got away from him in anger |
| Paco nos la sacó del barro | Paco took it out of the mud for us |
| Pedro nos las verá si puede | Peter will see them for us if he can |
| Sara se les llevó la lora | Sarah took the parrot away from them |
| Sergio me lo creyó del todo | Serge totally believed me |

Regular feet

| | |
|---|---|
| Él me la trae sin crédito | He brings it to me without credit |
| Él te la cambia por máquinas | He trades it with you for machines |
| Flor se te sale del médico | Flor gets away from the doctor's office |
| John me le compra la fórmula | John buys the formula for me |
| Juan se la sabe sin cálculos | John knows it without calculations |
| Luis me los paga con dólares | Lewis pays them to me with dollars |
| Luz nos los tiene sin pérdidas | Luz keeps them for us without losses |
| Mar me lo cuenta los miércoles | Mar tells it to me on Wednesdays |
| Juan te los llama con música | John calls them for you with music |
| Paz nos la debe del sábado | Paz owes us for Saturday |
| Paul se les ríe tras cámaras | Paul laughs at them behind the cameras |
| Rey nos las guarda con código | Ray keeps them for us with a code |
| Sol te lo lleva del público | Sol brings it to you from the public |
| Tú nos lo cuentas con método | You tell it to us with a method |
| Yo se lo digo sin lágrimas | I tell it to him without tears |
| Yo te las cambio por pájaros | I trade them with you for birds |

Irregular feet

| | |
|---|---|
| Él me siente de su público | He considers me a part of his audience |
| Él se ríe de mi síndrome | He laughs at my syndrome |
| Flor le debe por los músicos | Flor owes him for the musicians |
| John le paga sin tu fórmula | John pays him without your formula |
| Juan lo quiere sin los títulos | John wants it without the titles |
| Luis la lleva con sus líderes | Lewis brings her with his leaders |
| Luz lo guarda tras la cámara | Luz keeps it behind the camera |
| Mar nos cuenta de tus pérdidas | Mar tells us about your losses |
| Juan nos dijo de la práctica | John told us about the practice |
| Paz lo tiene con su plástico | Paz keeps it with its plastic |
| Paul me nota por mis músculos | Paul notices me for my muscles |
| Rey los compra con tus dólares | Ray buys them with your dollars |
| Sol se cae tras las máquinas | Sol falls down behind the machines |
| Tú los llamas por sus términos | You refer to them by their terms |
| Yo las cuento sin la técnica | I count them without the technique |
| Yo te veo tras mis lágrimas | I see you behind my tears |